

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-204442

(43)Date of publication of application : 05.08.1997

(51)Int.Cl.

G06F 17/30

(21)Application number : 08-031466

(71)Applicant : DAINIPPON SCREEN MFG CO LTD

(22)Date of filing : 24.01.1996

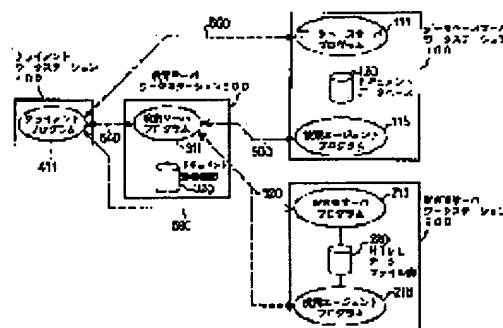
(72)Inventor : KURIHARA DAIKI
KADOMA HISAAKI

(54) DOCUMENT DATA RETRIEVAL SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To properly retrieve updated document data, to prevent unnecessary document data in a document server from being set to a retrieval object and to prevent the stagnation of a network and the increase of the loads of the document server and a retrieval server.

SOLUTION: When document data in a document data base 120 is updated in a data base server 100, a data base server program 111 generates a data update trigger and transmits it to a retrieval agent program 115. When the retrieval agent program 115 receives the trigger, it generates a document index based on updated document data and transfers it to a retrieval server 300. In the retrieval server 300, a retrieval server program 311 updates document index information 320 based on the transferred document index.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the
examiner's decision of rejection or application
converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-204442

(43) 公開日 平成9年(1997)8月5日

(51) Int.Cl.⁹
G 0 6 F 17/30

識別記号 庁内整理番号

F I
G 0 6 F 15/40
15/403

技術表示箇所

3 1 0 C
3 4 0 Z

審査請求 未請求 請求項の数 3 F D (全 16 頁)

(21) 出願番号 特願平8-31488

(22) 出願日 平成8年(1996)1月24日

(71) 出願人 000207551

大日本スクリーン製造株式会社
京都府京都市上京区堀川通寺之内上る4丁目天神北町1番地の1

(72) 発明者 栗原 大樹

京都市上京区堀川通寺之内上る4丁目天神北町1番地の1 大日本スクリーン製造株式会社内

(72) 発明者 角間 央章

京都市上京区堀川通寺之内上る4丁目天神北町1番地の1 大日本スクリーン製造株式会社内

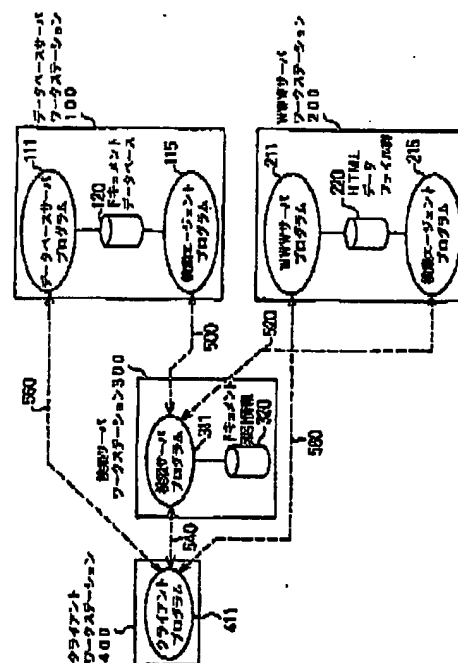
(74) 代理人 井理士 五十嵐 幸雄 (外3名)

(54) 【発明の名称】 ドキュメントデータ検索システム

(57) 【要約】

【課題】 更新されたドキュメントデータに対する検索を適正に行なうことができ、不要なドキュメントサーバ内のドキュメントデータを検索対象とすることなく、また、ネットワークの渋滞やドキュメントサーバ、検索サーバの負荷の増加を招くことのないようにする。

【解決手段】 データベースサーバ100において、ドキュメントデータベース120内のドキュメントデータが更新されたら、データベースサーバプログラム111はデータ更新トリガを発生して検索エージェントプログラム115に送信する。検索エージェントプログラム115はそのトリガを受信すると、更新されたドキュメントデータに基づいてドキュメント索引を生成し、検索サーバ300に転送する。検索サーバ300において、検索サーバプログラム311は、転送されたドキュメント索引に基づいてドキュメント索引情報320を更新する。



【特許請求の範囲】

【請求項1】 クライアントと、

複数のドキュメントデータを格納し、前記クライアントからの要求に応じて所望のドキュメントデータを提供する1つ以上のドキュメントサーバと、
各ドキュメントデータのサマリー情報によって構成されるドキュメント索引情報を格納し、前記クライアントからの要求に応じて、前記ドキュメント索引情報を参照しつつ、所望のドキュメントデータの格納場所を検索し、検索結果を提供する検索サーバと、
を通信回線を介して接続して成るドキュメントデータ検索システムであって、

前記ドキュメントサーバは、格納する前記ドキュメントデータが更新されたか否かを検出する検出手段と、
該検出手段によって前記ドキュメントデータの更新が検出されたら、更新された該ドキュメントデータから得られる該ドキュメントデータのサマリー情報に基づいて、更新情報を生成する生成手段と、
生成した前記更新情報を前記通信回線を介して前記検索サーバに転送する転送手段と、を備え、
前記検索サーバは、転送された前記更新情報に基づいて、格納する前記ドキュメント索引情報を更新する更新手段を備えるドキュメントデータ検索システム。

【請求項2】 請求項1に記載のドキュメントデータ検索システムにおいて、

前記ドキュメントサーバは、
格納している前記ドキュメントデータを、データベース管理システムによって管理する手段をさらに備えると共に、

前記検出手段は、前記ドキュメントデータの更新を検出すると、データ更新トリガを発生する手段を備え、

前記生成手段は、前記データ更新トリガの発生に応じて前記更新情報の生成を開始する手段を備えるドキュメントデータ検索システム。

【請求項3】 請求項1に記載のドキュメントデータ検索システムにおいて、

前記ドキュメントサーバは、
格納している前記ドキュメントデータを、オペレーティングシステムにおけるファイル管理システムによって管理する手段をさらに備えると共に、

前記検出手段は、管理しているファイルを順次チェックして、格納する前記ドキュメントデータが更新されたか否かを検出する手段を備えるドキュメントデータ検索システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、複数のドキュメントデータを1つ以上のドキュメントサーバに分散して格納している場合に、各ドキュメントサーバに通信回線を介して接続される検索サーバによって、所望のドキュメ

ントデータの格納場所をドキュメント索引情報に基づいて検索するドキュメントデータ検索システムに関するものである。

【0002】

【従来の技術】 ドキュメントデータを効率よく検索するためには、すべてのドキュメントデータを1つのドキュメントサーバに格納し、集中的に管理するのがよい。しかし、ドキュメントデータを扱う組織の規模が大きくなると、大量のドキュメントデータを扱うなければならなくなるため、1つのドキュメントサーバだけでは管理しきれず、それゆえ、大量のドキュメントデータを複数のドキュメントサーバなどに分散して格納する必要がある。

【0003】 このとき、或るドキュメントデータを必要とするユーザ（検索者）が、そのドキュメントデータがどのドキュメントサーバに格納されているかを知らない場合、検索者はクライアントを使って、書誌情報やキーワードなどに基づき、ドキュメントサーバ毎に、目的のドキュメントデータが格納されているか否かを検索しなければならない。これは検索者にとって大変負担が大きい。

【0004】 そこで、この問題を解決するために、ドキュメント索引情報を格納した検索サーバを用いた検索システムが提案されている。このような検索システムとしては、例えば、WWW (World Wide Web) により提供されるドキュメントデータを対象とした検索システムがある。WWWは、ハイパーテキスト形式のドキュメントデータを提供する仕組みである。WWWでは、このようなハイパーテキスト形式のドキュメントデータをWWWサーバに格納しており、一つのドキュメントデータから他のドキュメントデータへはリンクを張ることができる。検索者は、そのリンクをたどることによって、目的とするドキュメントデータを得る。しかし、WWWでは、このようなリンクをたどること以外には、ドキュメントデータを検索する手段が提供されていない。そこで、これを補うために、前述した検索サーバを用いた検索システムが種々提案されている。例えば、「情報処理学会、第2回、テクニカルコミュニケーションシンポジウム予備集」（'95.7）の第24頁～第31頁に掲載されている「SGML文書管理システム実現における課題」

（井上直樹：NTTデータ通信株式会社）や、或いは「情報処理学会、第51回、全国大会」（'95.9）の論文集の1-169～170に掲載されている「WWWにおける広域検索システム」（田村健人、村岡洋一：早稲田大学理工学部）においては、WWWを対象とした実際の検索サーバの例が開示されている。

【0005】 この検索サーバは、ネットワーク上の全てのWWWサーバに対して、自ら定期的にアクセスして、各WWWサーバに格納されている全ドキュメントデータ（全ページの全テキスト）をそれぞれ取得して、その取

3

得したデータを基にドキュメント索引情報を生成し、そのドキュメント索引情報を格納している。そして、検索者がクライアントを使って、その検索サーバにアクセスすると、検索サーバは格納しているドキュメント索引情報を用いて、目的とするドキュメントデータの格納場所を検索し、その検索結果を検索者に知らせる。

【0006】

【発明が解決しようとする課題】このような従来における検索サーバを用いた検索システムにおいては、次のような問題があった。

【0007】①或るWWWサーバ内のドキュメントデータが更新されても、その後、検索サーバが、そのWWWサーバに対してアクセスしてその更新されたドキュメントデータを取得した上で、それらデータに基づいて検索サーバ内のドキュメント索引情報が更新されるまでの間は、上記ドキュメントデータの更新がドキュメント索引情報に反映されないため、そのドキュメントデータに対する検索が適正に行なわれない。

【0008】②検索サーバは、ネットワーク上の全てのWWWサーバからデータを収集してドキュメント索引情報を生成しているため、不要なWWWサーバ内のドキュメントデータまでもが、検索対象となる可能性がある。

【0009】③検索サーバは、WWWサーバに定期的にアクセスして全ドキュメントデータを取得するようにしているため、そのWWWサーバ内のドキュメントデータに一切更新が生じていない場合には、そのようなアクセスやデータの取得は無駄であって、ネットワークの渋滞やWWWサーバへの負荷の増加を招く原因となりがねない。

【0010】従って、本発明の目的は、上記した従来技術の問題点を解決し、更新されたドキュメントデータに対する検索を適正に行なうことができ、不要なドキュメントサーバ内のドキュメントデータを検索対象とすることなく、また、ネットワークの渋滞やドキュメントサーバ、検索サーバの負荷の増加を招くことのないドキュメントデータ検索システムを提供することにある。

【0011】

【課題を解決するための手段およびその作用・効果】上記した目的の少なくとも一部を達成するために、第1の発明は、クライアントと、複数のドキュメントデータを格納し、前記クライアントからの要求に応じて所望のドキュメントデータを提供する1つ以上のドキュメントサーバと、各ドキュメントデータのサマリー情報によって構成されるドキュメント索引情報を格納し、前記クライアントからの要求に応じて、前記ドキュメント索引情報を参照しつつ、所望のドキュメントデータの格納場所を検索し、検索結果を提供する検索サーバと、を通信回線を介して接続して成るドキュメントデータ検索システムであって、前記ドキュメントサーバは、格納する前記ドキュメントデータが更新されたか否かを検出する検出手

4

段と、該検出手段によって前記ドキュメントデータの更新が検出されたら、更新された該ドキュメントデータから得られる該ドキュメントデータのサマリー情報に基づいて、更新情報を生成する生成手段と、生成した前記更新情報を前記通信回線を介して前記検索サーバに転送する転送手段と、を備え、前記検索サーバは、転送された前記更新情報に基づいて、格納する前記ドキュメント索引情報を更新する更新手段を備えることを要旨とする。

【0012】ここで、ドキュメントデータとしては、テキストデータ（HTMLデータなども含む）や、画像データや音声データなどのバイナリデータなどが挙げられる。また、ドキュメントデータのサマリー情報としては、ドキュメント名や作者名などの番誌情報や、キーワードや、ドキュメント種別や、ドキュメントデータの格納場所などが挙げられる。

【0013】このように、本発明では、ドキュメントサーバ側において、検出手段が、格納するドキュメントデータの更新を検出したら、生成手段が、更新されたドキュメントデータから得られるサマリー情報に基づいて更新情報を生成し、転送手段が、その更新情報を通信回線を介して検索サーバに転送する。一方、検索サーバ側においては、更新手段が、転送された更新情報に基づいて、格納するドキュメント索引情報を更新する。

【0014】従って、本発明によれば、以下に掲げる効果を奏する。

①ドキュメントサーバに格納されたドキュメントデータが更新されても、その更新情報はすぐに検索サーバに転送され、それに基づいて、格納するドキュメント索引情報が更新されるので、ドキュメントデータの更新は直ちにドキュメント索引情報に反映され、そのドキュメント索引情報を基にして、その更新されたドキュメントデータに対する検索を適正に行なうことができる。

【0015】②検索サーバは、ネットワーク上の全てのドキュメントサーバを相手とするのではなく、上記した検出手段、生成手段及び転送手段を備えたドキュメントサーバのみを相手として、ドキュメント索引情報を得ているため、不要なドキュメントサーバ内のドキュメントデータは一切検索対象とならない。

【0016】③ドキュメントサーバは、格納しているドキュメントデータが更新された場合、格納する全てのドキュメントデータについて更新情報を生成するのではなく、更新されたドキュメントデータについてのみ更新情報を生成して、検索サーバに転送するため、ネットワークの渋滞やドキュメントサーバ、検索サーバの負荷の増加を招く恐れがない。

【0017】本発明のドキュメントデータ検索システムにおいて、前記ドキュメントサーバは、格納している前記ドキュメントデータを、データベース管理システムによって管理する手段をさらに備えると共に、前記検出手段は、前記ドキュメントデータの更新を検出すると、デ

5

ータ更新トリガを発生する手段を備え、前記生成手段は、前記データ更新トリガの発生に応じて前記更新情報の生成を開始する手段を備えるようにしても良い。

【0018】このように、ドキュメントサーバにおいて、格納するドキュメントデータをデータベース管理システムによって管理する場合、データ更新トリガを発生する手段を備えることができる。即ち、この発生手段は、ドキュメントデータの更新があった場合に、データ更新トリガを発生するため、生成手段では、これを利用して、更新情報の生成を開始させることができる。

【0019】また、本発明のドキュメントデータ検索システムにおいて、前記ドキュメントサーバは、格納している前記ドキュメントデータを、オペレーティングシステムにおけるファイル管理システムによって管理する手段をさらに備えると共に、前記検出手段は、管理しているファイルを順次チェックして、格納する前記ドキュメントデータが更新されたか否かを検出する手段を備えるようにしても良い。

【0020】このように、ドキュメントサーバにおいて、格納するドキュメントデータをオペレーティングシステムにおけるファイル管理システムによって管理する場合、データベース管理システムによって管理する場合のように、データ更新トリガを発生する手段を備えることはできない。そのため、管理しているファイルを順次チェックして更新の有無を検出する手段を備えるようにする。このような手段を用いれば、ファイル管理されているドキュメントデータについても、漏れなくドキュメントデータの更新を検出することができる。

【0021】

【発明の実施の形態】以下、本発明の実施の形態を実施例に基づいて説明する。図1は本発明の実施例としてドキュメントデータ検索システムの概要を示す説明図であり、図2は図1のドキュメントデータ検索システムの詳細な構成を示すブロック図である。

【0022】図1または図2に示すように、このドキュメントデータ検索システムは、データベースサーバワークステーション100と、WWWサーバワークステーション200と、検索サーバワークステーション300と、クライアントワークステーション400と、を備えており、それらは互いにネットワーク上の通信回線500、520、540、560、580によって接続されている。

【0023】データベースサーバワークステーション（以下、データベースサーバと略す。）100は、図2に示すように、各種プログラムを記憶するプログラムメモリ110と、複数のドキュメントデータによって構成されるドキュメントデータベース120と、プログラムメモリ110内のプログラムに従って各種処理動作を行なうCPU130と、ネットワークを介して他のワークステーションと通信を行なうためのネットワークインタ

6

フェース140と、更新レコードを一時的に記憶する更新バッファ150を備えている。ここで、ドキュメントデータベース120を構成する各ドキュメントデータは、データベース管理システム(Data BaseManagement System)によって管理されている。

【0024】プログラムメモリ110には、プログラムとして、ドキュメントデータを参照したり、更新したりするためのデータベースサーバプログラム111や、ドキュメント索引情報を生成し、転送するための検索エージェントプログラム115が記憶されている。

【0025】図3はこのようなデータベースサーバプログラム111及び検索エージェントプログラム115による機能を説明するための説明図である。CPU130は、データベースサーバプログラム111に従って処理動作をすることによって、図3(a)に示すように、ドキュメントデータ参照手段112や、ドキュメントデータ更新手段113や、データ更新トリガ送信手段114として機能し、また、検索エージェントプログラム115に従って処理動作をすることによって、図3(b)に示すように、データ更新トリガ受信手段116や、ドキュメント索引生成手段117と、ドキュメント索引転送手段118として機能する。

【0026】また、WWWサーバワークステーション（以下、WWWサーバと略す。）200は、図2に示すように、プログラムメモリ210と、複数のHTML(HyperText Mark-up Language)データによって構成されるHTMLデータファイル群220と、CPU230と、ネットワークインタフェース240と、バッファ群250を備えている。バッファ群250は、ファイル一覧バッファ252と、前回ファイル一覧バッファ254と、更新バッファ256を備えている。ここで、各HTMLデータは、オペレーティングシステムにおけるファイル管理システムによって管理されている。

【0027】プログラムメモリ210には、プログラムとして、HTMLデータを参照するためのWWWサーバプログラム211と、HTMLデータの更新を検知してドキュメント索引情報を生成し、転送するための検索エージェントプログラム215が記憶されている。

【0028】図4はこのようなWWWサーバプログラム211及び検索エージェントプログラム215による機能を説明するための説明図である。CPU230は、WWWサーバプログラム211に従って処理動作をすることによって、図4(a)に示すように、HTMLデータ参照手段212として機能し、また、検索エージェントプログラム215に従って処理動作をすることによって、図4(b)に示すように、データ更新検知手段216や、ドキュメント索引生成手段217や、ドキュメント索引転送手段218として機能する。

【0029】また、検索サーバワークステーション（以下、検索サーバと略す。）300は、図2に示すよう

に、プログラムメモリ310と、ドキュメント索引情報320と、CPU330と、ネットワークインタフェース340を備えている。

【0030】プログラムメモリ310には、プログラムとして、ドキュメント索引情報320を検索したり、更新したりするための検索サーバプログラム311が記憶されている。

【0031】図5はこのような検索サーバプログラム311による機能を説明するための説明図である。CPU330は、検索サーバプログラム311に従って処理動作をすることによって、図5に示すように、ドキュメント索引検索手段312や、ドキュメント索引更新手段313として機能する。

【0032】クライアントワークステーション（以下、クライアントと略す。）400は、図2に示すように、プログラムメモリ410と、CPU430と、ネットワークインタフェース440を備えている。プログラムメモリ410には、プログラムとして、データベースサーバ100やWWWサーバ200或いは検索サーバ300にアクセスするためのクライアントプログラム411が記憶されている。

【0033】さて、検索サーバ300が初めて運用する場合、ドキュメント索引情報320は無い状態であるので、データベースサーバ100やWWWサーバ200では、それぞれ、検索エージェントプログラム115、215が起動し、格納している全ドキュメントデータについて、それぞれ、ドキュメントデータの書誌情報やキーワードやデータ格納場所などを含むサマリー情報（ドキュメント索引）を生成し、検索サーバ300に転送する。検索サーバ300では、検索サーバプログラム311が起動して、転送された複数のドキュメント索引を受け取って、ドキュメント索引情報320として格納する。

【0034】次に、データベースサーバ100において、ドキュメントデータベース120内のドキュメントデータを更新（即ち、追加、変更、削除）する場合の動作について説明する。

【0035】図6は図2のデータベースサーバプログラム111の処理内容を示すフローチャート、図7は図2のドキュメントデータベース120内のドキュメントデータを更新する場合の一例を示すフローチャート、図8は図2のドキュメントデータベース120の内容と更新バッファ150の内容の一例を示す説明図、図9はドキュメントデータのデータ構造の一例を示す説明図である。

【0036】ドキュメントデータベース120は、例えば、複数のドキュメントデータファイルによって構成され、さらに、各ドキュメントデータファイルは図8

(a)に示すように複数のドキュメントデータによって構成されている。また、各ドキュメントデータは、それ

ぞれ、図9に示すように、本文の他、ドキュメント名や、作者名などによって構成されている。

【0037】さて、図2に示すデータベースサーバプログラム111が起動されると、図6に示すように、まず、コマンド待ちの状態となり（ステップS102）、その後、コマンドが入力されると、入力されたコマンドに応じて処理が分かれる（ステップS104）。

【0038】ドキュメントデータベース120内のドキュメントデータを更新する場合、その更新はトランザクション単位毎に行なわれる。例えば、図7に示すように、トランザクションが開始されると（ステップS202）、データベースサーバプログラム111には、図6に示すように、コマンドとしてトランザクション開始が入力されるので、ステップS124に進んで更新行数を0にすると共に、更新バッファ150をクリアする（ステップS126）。

【0039】次に、図8(a)に示すドキュメントデータベース120に格納されたファイル1について、図7に示すように、ドキュメントデータ1を追加する場合（ステップS204）、データベースサーバプログラム111には、図6に示すように、コマンドとしてデータ追加が入力されるので、ドキュメントデータベース120にドキュメントデータ1を1行追加し（ステップS110）、更新バッファ150にも、図8(b)に示すようにその追加行（即ち、ドキュメントデータ1）をその種別（この場合、「追加」と共に追加して（ステップS112）、更新レコード151とする。そして、更新行数に1を加える（ステップS122）。

【0040】次に、図7に示すように、ドキュメントデータ2を変更する場合（ステップS206）、データベースサーバプログラム111には、図6に示すように、コマンドとしてデータ変更が入力されるので、ドキュメントデータベース120の該当行のドキュメントデータ2を変更し（ステップS114）、更新バッファ150にも、図8(b)に示すようにその変更行（即ち、ドキュメントデータ2）をその種別（この場合、「変更」と共に追加して（ステップS116）、更新レコード152とする。そして、更新行数に1を加える（ステップS122）。

【0041】さらに、図7に示すように、ドキュメントデータ3を削除する場合（ステップS208）、データベースサーバプログラム111には、図6に示すように、コマンドとしてデータ削除が入力されるので、ドキュメントデータベース120から該当行のドキュメントデータ3を削除し（ステップS118）、更新バッファ150にも、図8(b)に示すようにその削除行（即ち、ドキュメントデータ3）をその種別（この場合、「削除」と共に追加して（ステップS116）、更新レコード153とする。そして、更新行数に1を加える（ステップS122）。

【0042】そして、図7に示すように、トランザクションが終了すると（ステップS210）、データベースサーバプログラム111には、図6に示すように、コマンドとしてトランザクション終了が入力されるので、データ更新トリガを発生して検索エージェントプログラム115に送る（ステップS128）。

【0043】以上のようなドキュメントデータの更新とデータ更新トリガの発生・送信は図3（a）に示すドキュメントデータ更新手段113とデータ更新トリガ送信手段114がそれぞれ行なうことになる。なお、図6に

10 において、データ参照については、後ほど説明する。
【0044】以上のように、ドキュメントデータベース120内の1トランザクション分のドキュメントデータの更新が終了すると、更新バッファ150には、図8（b）に示すように、更新されたドキュメントデータとその種別（「追加」、「変更」、「削除」）から成る更新レコード151～153が1トランザクション分記憶された状態となる。

【0045】次に、データベースサーバ100において、更新されたドキュメントデータに基づいてドキュメント索引を生成する場合の動作について説明する。

【0046】図10は図2の検索エージェントプログラム115の処理内容を示すフローチャート、図11は通信回線500を介してデータベースサーバ100から検索サーバ300へ転送される転送データの一例を示す説明図、図12はドキュメント索引のデータ構造の一例を示す説明図である。

【0047】図2に示す検索エージェントプログラム115が起動されると、図10に示すように、まず、イベント待ちの状態となり（ステップS302）、その後、イベントが入力されると、そのイベントがデータ更新トリガイイベントか否かが判定される（ステップS304）。即ち、データベースサーバプログラム111から送信されたデータ更新トリガを受信したか否かが判定される。判定の結果、データ更新トリガイイベントでなく、別のイベントであれば、ステップS314に進んでそのイベントに応じた処理を行なう。

【0048】一方、データ更新トリガイイベントであれば、図8（b）に示す更新バッファ150より更新レコードの一つを取り出す（ステップS306）。そして、その更新レコードのドキュメントデータに基づいてドキュメント索引を生成し、種別と共に転送レコードを構成させる（ステップS308）。例えば、取り出した更新レコードが図8（b）に示す更新レコード151であるとすると、その更新レコード151内のドキュメントデータ1からドキュメント索引1を生成し、更新レコード151内の種別（この場合、「追加」）と共に、図11に示す転送レコード504を構成させる。

【0049】ここで、生成されるドキュメント索引は、図12に示すように、ドキュメント名や作者名などの資

誌情報の他、キーワードや、ドキュメント種別や、ドキュメントデータの格納場所などによって構成される。このうち、番誌情報やキーワードは、図9に示したドキュメントデータの中から、それぞれ抽出される。また、ドキュメント種別としては、テキストデータ、バイナリデータの別や、さらに細かく画像データや、HTMLデータの別などが付される。また、ドキュメントデータの格納場所としては、データベースサーバ内のドキュメントデータベースに格納されたドキュメントデータの場合、データベースサーバ名、データベース名、ファイル名、主キー値（データベース内のドキュメントを一意的に区別するための識別番号のようなもの）などが付される。また、後述するようなWWWサーバ内のファイルに格納されたHTMLデータの場合は、情報資源のアクセス方法と存在場所を指定するためのURL（Uniform Resource Locator）などが付される。

【0050】なお、更新レコード内の種別が「変更」や「削除」である場合、生成するドキュメント索引としては、図12に示すような情報を全て有する必要はない。即ち、「変更」の場合は少なくとも変更された部分の情報（例えば、作者名が変更されている場合は、その作者名）とドキュメントデータの格納場所を有していれば良い。また、「削除」の場合は少なくともドキュメントデータの格納場所を有していれば良い。

【0051】また、ドキュメントデータの種別によっては、図12に示す情報の全てが得られるとは限らないし、異なる情報が得られる場合もある。従って、そのような場合には、得られる情報によってドキュメント索引を構成すればよい。

【0052】次に、ステップS306で取り出した更新レコードが最後のレコードであるか否かを判定する（ステップS310）。判定の結果、最後のレコードでなければ、ステップS306に戻って、更新バッファ150より次の更新レコードの一つを取り出し、同様の処理を繰り返す。最後のレコードであれば、1トランザクション分の更新レコードについて、ドキュメント索引生成の処理が終了したことになるので、図11に示すように1トランザクション分の転送レコード504～508が得られたことになる。

【0053】次に、こうして得られた1トランザクション分の転送レコード504～508をドキュメント索引更新要求502と共に、通信回線500を介して検索サーバ300の検索サーバプログラム311に転送する（ステップS312）。即ち、検索サーバプログラム311へは、図11に示すように、先頭から、ドキュメント索引更新要求502、転送レコード504～508の順に順次転送される。

【0054】以上のようなデータ更新トリガの受信とドキュメント索引の生成とドキュメント索引の転送は、図3（b）に示すデータ更新トリガ受信手段116とドク

11

ュメント索引生成手段117とドキュメント索引転送手段118がそれぞれ行なうことになる。

【0055】次に、検索サーバ300において、転送されたドキュメント索引を基に、格納されているドキュメント索引情報を更新する場合の動作について説明する。

【0056】図13は図2の検索サーバプログラム311の処理内容の一部を示すフローチャート、図14は図2の検索サーバプログラム311の処理内容の残りの部分を示すフローチャート、図15は図2のドキュメント索引情報320の内容の一例を示す説明図である。

【0057】図2に示す検索サーバプログラム311が起動されると、図13に示すように、まず、イベント待ちの状態となり（ステップS402）、その後、イベントが入力されると、そのイベントが検索エージェントプログラムからのイベントかクライアントプログラムからのイベントかが判定される（ステップS404）。判定の結果、クライアントプログラムからのイベントである場合には、図14のAに進む。

【0058】一方、検索エージェントプログラムからのイベントである場合には、さらに、そのイベントがドキュメント索引更新要求イベントであるか否かが判定される（ステップS408）。即ち、検索エージェントプログラムから転送されてきた図11に示すドキュメント索引更新要求502を受信したか否かが判定される。判定の結果、ドキュメント索引更新要求イベントでなく、別のイベントであれば、ステップS422に進んで、そのイベントに応じた処理を行なう。

【0059】ドキュメント索引更新要求イベントであれば、図11に示す次に転送されてくる転送レコードを1つ受信する（ステップS410）。そして、その転送レコードの種別が何かを判定し（ステップS412）、その判定結果に応じた処理を行なう。例えば、図11に示すように、ドキュメント索引更新要求502に続く転送レコードが504である場合、その種別は「追加」であるので、ステップS414に進み、図15に示すように、ドキュメント索引情報320の最後尾に、転送レコード504内のドキュメント索引1を1行追加する。

【0060】次に、ステップS410で受信した転送レコードが最後のレコードであるか否かを判定する（ステップS420）。判定の結果、最後のレコードでなければ、ステップS410に戻って、次に転送されてくる転送レコードを1つ受信する。図11に示すように、次に続く転送レコードは506であり、その種別は「変更」であるので、ステップS416に進み、図15に示すように、ドキュメント索引情報320内の該当行のドキュメント索引2を、転送レコード内の新たなドキュメント索引2に変更する。このとき、ドキュメント索引情報320内の該当行のドキュメント索引は、ドキュメント索引情報320内の複数のドキュメント索引の中から、ドキュメントデータの格納場所が、転送レコード内のドク

12

ュメント索引と一致するものを探し出すことによって、容易に見つけ出すことができる。

【0061】さらに、ステップS410に戻って、次に転送されてくる転送レコードを1つ受信した場合、図11に示すように、次に続く転送レコードは508であり、その種別は「削除」であるので、ステップS418に進み、図15に示すように、ドキュメント索引情報320内の該当行のドキュメント索引3を削除する。

【0062】一方、ステップS420において、最後のレコードであると判定された場合には、検索エージェントプログラムから転送されてきた全ての転送レコードを受信したことになるので、ドキュメント索引情報320の更新処理を終了して、再びイベント待ち（ステップS402）の状態に戻る。以上のようなドキュメント索引情報320の更新は図5に示すドキュメント索引更新手段313が行なうことになる。

【0063】次に、検索者が目的とするドキュメントデータを検索し、そのドキュメントデータを得る場合の動作について説明する。まず、クライアント400において、クライアントプログラム411を起動する。そして、検索者が、検索したいドキュメントデータのドキュメント名、作者名などの書誌情報の一部やキーワードなどを、検索条件としてクライアント400に入力すると、クライアントプログラム411は、検索要求と検索条件を、通信回線540を介して検索サーバ300の検索サーバプログラム311に転送する。

【0064】一方、検索サーバプログラム311では、図13において既に説明したように、イベントが入力されると、そのイベントが検索エージェントプログラムからのイベントかクライアントプログラムからのイベントかを判定し（ステップS404）、クライアントプログラムからのイベントである場合には、図14のAに進む。図14では、まず、そのイベントが検索要求イベントであるか否かを判定する（ステップS424）。即ち、クライアントプログラムから転送されてきた検索要求を受信したか否かを判定する。判定の結果、検索要求イベントでなく、別のイベントであれば、ステップS432に進んで、そのイベントに応じた処理を行なう。

【0065】検索要求イベントであれば、次に転送されてくる検索条件を受信し（ステップS426）、格納するドキュメント索引情報320からその検索条件を満たす行、即ち、ドキュメント索引を抽出する（ステップS428）。そして、その抽出結果を通信回線540を介してクライアント400のクライアントプログラム411に送信する（ステップS430）。このようなドキュメント索引情報320に対する検索は図5に示すドキュメント索引検索手段312が行なうことになる。

【0066】クライアントプログラム411は、送信された検索結果を受信すると、それを検索者に対して提示し、それにより、検索者は目的とするドキュメントデー

タの格納場所や、その他必要な情報を知ることができる。

【0067】次に、検索者がそのドキュメントデータの取得をクライアントプログラム411に要求すると、クライアントプログラム411は、そのドキュメントデータの格納されているサーバにアクセスを開始する。例えば、そのサーバがデータベースサーバ100である場合、クライアントプログラム411は通信回線560を介してデータベースサーバ100にアクセスして、データベースサーバプログラム111に対しデータ参照コマンドを送出する。

【0068】クライアントプログラム411は、図6に示したように、ステップS104において、入力されたコマンドがデータ参照コマンドであると判定すると、ステップS106に進んで、ドキュメントデータベース120の中から、該当行のドキュメントデータを読み出して（ステップS106）、読み出したその行、即ち、目的とするドキュメントデータを出力媒体（この場合、クライアント400）に対して転送する。このような目的とするドキュメントデータの読み出しは、図3(a)のドキュメントデータ参照手段112によって行なわれる。

【0069】このようにして転送されたドキュメントデータを、クライアント400のクライアントプログラム411が受信することによって、検索者は目的とするドキュメントデータを得ることができる。

【0070】次に、WWWサーバ200において、HTMLデータファイル群220内のHTMLデータを更新（即ち、追加、変更、削除）した場合の動作について説明する。

【0071】前述したように、データベースサーバ100では、格納しているドキュメントデータの管理をデータベース管理システムによって行なっているが、WWWサーバ200では、格納しているHTMLデータの管理を、オペレーティングシステムにおけるファイル管理システムによって行なっている。そのため、WWWサーバ200のWWWサーバプログラム211には、図4

(a)に示すように、データベースサーバ100のデータベースサーバプログラム111のようなデータ更新トリガ送信手段114は存在しない。そのため、HTMLデータファイル群220内のHTMLデータが更新（即ち、追加、変更、削除）されても、WWWサーバプログラム211から検索エージェントプログラム215へはデータ更新トリガが送られないため、検索エージェントプログラム215では、HTMLデータが更新されたかどうかを知ることができない。そこで、本実施例においては、図4(b)に示すように検索エージェントプログラム215にデータ更新検知手段216を設けている。

【0072】図16は図2の検索エージェントプログラム215におけるデータ更新検知手段216の処理内容

の前半部分を示すフローチャート、図17は図2の検索エージェントプログラム215におけるデータ更新検知手段216の処理内容の後半部分を示すフローチャート、図18は図2のファイル一覧バッファ252、前回ファイル一覧バッファ254、更新バッファ256の記憶内容の一例を示す説明図である。

【0073】図16に示すように、処理が開始されると、まず、一定時間スリープした後（ステップS502）、HTMLデータファイル群220内のHTMLデータの格納されているファイルの一つを選択する（ステップS504）。即ち、HTMLデータファイル群220では、各ファイルはそれぞれディレクトリで階層に分類されて收容されており、ここでは、WWWサーバプログラム211が管理する特定のディレクトリ以下のHTMLデータの格納されているファイルが全て選択の対象となる。

【0074】なお、この段階において、図18に示す3つのバッファのうち、ファイル一覧バッファ252と更新バッファ256については記憶内容が全てクリアされているが、前回ファイル一覧バッファ254については、前回格納したファイルのパス名やタイムスタンプがそのままとなっており、「処理済み」の欄のみがクリアされているものとする。

【0075】次に、選択したファイルから、そのファイルのパス名とタイムスタンプを取り込んで、図18

(a)に示すファイル一覧バッファ252にそれぞれ格納する（ステップS506）。そして、取得したパス名及びタイムスタンプを、図18(b)に示す前回ファイル一覧バッファ254内に格納されている複数のパス名及びタイムスタンプと順次比較して、前回ファイル一覧バッファ254内に、取得したパス名とタイムスタンプの両方と一致するパス名及びタイムスタンプがあるかを判定する（ステップS508）。一致するパス名及びタイムスタンプがある場合にはステップS516に進むが、ない場合には、今度は、前回ファイル一覧バッファ254内に、取得したパス名と一致するパス名があるかを判定する（ステップS510）。一致するパス名がある場合にはステップS514に進み、ない場合にはステップS512に進む。

【0076】即ち、ステップS512に進んだ場合は、前回ファイル一覧バッファ254内に、取得したパス名と同じものがない場合であって、これは、ステップS504で選択したファイルが新たに追加されたファイルであることを示している。従って、ステップS512では、図18(c)に示す更新バッファ256に、取得したパス名と、種別として、追加されたファイルであることを示す「追加」を格納する。

【0077】また、ステップS514に進んだ場合は、前回ファイル一覧バッファ254内に、取得したパス名と同じものはあるが、タイムスタンプは異なっている場

合であって、これは、選択したファイルが前回のファイルに変更を加えたファイルであることを示している。従って、ステップS514では、図18(c)に示す更新バッファ256に、取得したパス名と、種別として、変更されたファイルであることを示す「変更」を格納する。そして、前回ファイル一覧バッファ254内の一一致したパス名について、「処理済み」の欄にチェックを付す(ステップS516)。

【0078】一方、ステップS508からステップS516に進んだ場合は、前回ファイル一覧バッファ254内に、取得したパス名と同じものがあり、タイムスタンプも同じである場合であって、これは、選択したファイルが、前回のファイルと全く同じであり、何ら変更が加えられていないことを示している。従って、この場合は、前回ファイル一覧バッファ254内の一一致したパス名について、「処理済み」の欄にチェックを付す(ステップS516)ことだけを行なう。

【0079】次に、ステップS518では、ステップS504で選択したファイルが、選択対象となっているファイルの中で、最後のファイルであるか否か判定し、最後のファイルでなければ、ステップS504に戻って、次のファイルを1つ選択し、以下同様の処理を行なう。最後のファイルであれば、図17のCに進む。

【0080】なお、前回ファイル一覧バッファ254内にパス名がある場合であっても、そのパス名に該当するファイルが、選択対象となっているファイルの中にある場合がある。即ち、前回存在していたファイルが削除された場合である。そのような削除されたファイルは図16に示した処理の中では見出すことができない。

【0081】そこで、図17では、まず、前回ファイル一覧バッファ254内において、「処理済み」の欄がチェックされていないパス名を探し出す。即ち、前回存在していたが、その後、削除されたファイルは、図16に示した処理がなされないため、前回ファイル一覧バッファ254内のそのファイルに該当するパス名について、「処理済み」の欄にチェックが付されないからである。従って、そのようなパス名を探し出したら、図18

(c)に示す更新バッファ256に、そのパス名と、種別として、削除されたファイルであることを示す「削除」を格納する(ステップS520)。

【0082】次に、更新バッファ256に記憶されている内容(即ち、パス名と種別)を図4(b)に示すドキュメント索引生成手段217に転送する(ステップS522)。そして、更新バッファ256と前回ファイル一覧バッファ254の記憶内容をクリアした後(ステップS524)、ステップS506で格納したファイル一覧バッファ252の記憶内容を前回ファイル一覧バッファ254に格納し(ステップS526)、その後、ファイル一覧バッファ252の記憶内容をクリアする(ステップS528)。以上の処理が終了したら、図16のDに

進んで、再び、一定時間のスリープ状態(ステップS502)に戻る。こうして、検索エージェントプログラム215におけるデータ変更検知手段216によって、HTMLデータファイル群220内におけるHTMLデータの更新の有無を検知することができる。

【0083】なお、検索エージェントプログラム215(即ち、ドキュメント索引生成手段217、ドキュメント索引転送手段218)によるドキュメント索引の生成やドキュメント索引の転送は、データベースサーバ100における検索エージェントプログラム115と処理内容はほぼ同じなので、説明は省略する。

【0084】以上のように、本実施例によれば、データベースサーバ100やWWWサーバ200に格納されたドキュメントデータ(HTMLデータも含む)が更新されても、その更新情報はドキュメント索引としてすぐに検索サーバ300に転送され、それに基づいて、格納するドキュメント索引情報320が更新されるので、ドキュメントデータの更新は直ちにドキュメント索引情報320に反映され、そのドキュメント索引情報320を基にして、その更新されたドキュメントデータに対する検索を適正に行なうことができる。

【0085】また、検索サーバ300は、ネットワーク上の全てのドキュメントサーバを相手とするのではなく、検索エージェントプログラムの稼働するドキュメントサーバ(即ち、データベースサーバ100及びWWWサーバ200)のみを相手として、ドキュメント索引情報320を得ているため、不要なドキュメントサーバ内のドキュメントデータは一切検索対象とならない。

【0086】さらにまた、データベースサーバ100及びWWWサーバ200は、格納しているドキュメントデータが更新された場合、格納する全てのドキュメントデータについてドキュメント索引を生成するのではなく、更新されたドキュメントデータについてのみ、ドキュメント索引を生成して、検索サーバ300に転送するため、ネットワークの渋滞やデータベースサーバ100、WWWサーバ200及び検索サーバ300の負荷の増加を最小限におさえることができる。

【0087】なお、本発明は上記した実施例や実施形態に限られるものではなく、その要旨を逸脱しない範囲において種々の態様にて実施することが可能である。

【0088】上記した実施例においては、ネットワークに接続されているドキュメントサーバは、データベースサーバ100とWWWサーバ200の2つであったが、1つであっても良いし、3つ以上であっても良い。同じく、クライアントもネットワーク上にクライアント400が1つあるだけであり、検索サーバも検索サーバ300が1つだけであったが、これらについても、2つ以上あって差し支えない。

【0089】また、上記した実施例では、検索サーバ300はドキュメントサーバと独立した構成となっている

が、一つのドキュメントサーバの中に、検索サーバの機能を持たせて、そのドキュメントサーバに検索サーバを兼ねさせるようにしても良い。

【0090】また、上記した実施例では、オペレーティングシステムにおけるファイル管理システムによって、HTMLデータを管理している例を挙げたが、HTMLデータに限らず、他のドキュメントデータを管理するようにしても良く、その場合にも、図16に示したような検索エージェントプログラム215におけるデータ変更検知手段216による処理を適用することができる。

【図面の簡単な説明】

【図1】本発明の一実施例としてドキュメントデータ検索システムの概要を示す説明図である。

【図2】図1のドキュメントデータ検索システムの詳細な構成を示すブロック図である。

【図3】図2のデータベースサーバプログラム111及び検索エージェントプログラム115による機能を説明するための説明図である。

【図4】図2のWWWサーバプログラム211及び検索エージェントプログラム215による機能を説明するための説明図である。

【図5】図2の検索サーバプログラム311による機能を説明するための説明図である。

【図6】図2のデータベースサーバプログラム111の処理内容を示すフローチャートである。

【図7】図2のドキュメントデータベース120内のドキュメントデータを更新する場合の一例を示すフローチャートである。

【図8】図2のドキュメントデータベース120の内容と更新バッファ150の内容の一例を示す説明図である。

【図9】ドキュメントデータのデータ構造の一例を示す説明図である。

【図10】図2の検索エージェントプログラム115の処理内容を示すフローチャートである。

【図11】通信回線500を介してデータベースサーバ100から検索サーバ300へ転送される転送データの一例を示す説明図である。

【図12】ドキュメント索引のデータ構造の一例を示す説明図である。

【図13】図2の検索サーバプログラム311の処理内容の一部を示すフローチャートである。

【図14】図2の検索サーバプログラム311の処理内容の残りの部分を示すフローチャートである。

【図15】図2のドキュメント索引情報320の内容の一例を示す説明図である。

【図16】図2の検索エージェントプログラム215におけるデータ更新検知手段216の処理内容の前半部分を示すフローチャートである。

【図17】図2の検索エージェントプログラム215に

におけるデータ更新検知手段216の処理内容の後半部分を示すフローチャートである。

【図18】図2のファイル一覧バッファ252、前回ファイル一覧バッファ254、更新バッファ256の記憶内容の一例を示す説明図である。

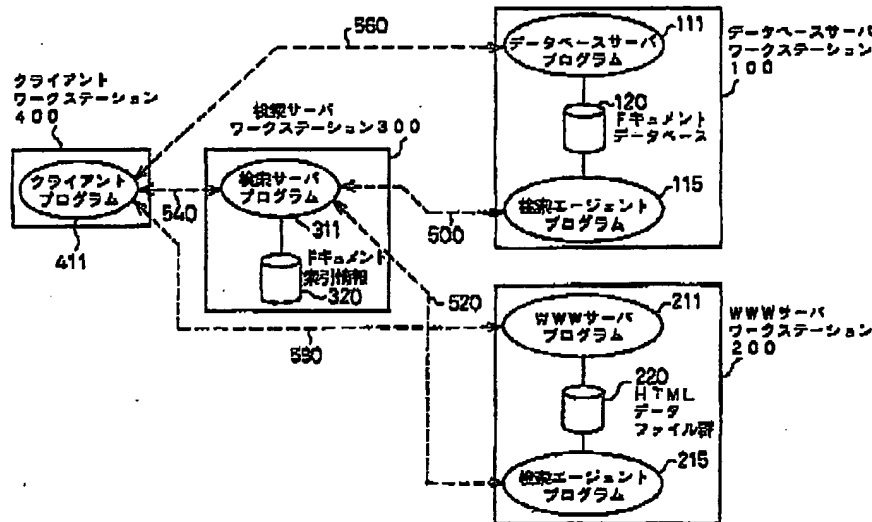
【符号の説明】

100…データベースサーバ
110…プログラムメモリ
111…データベースサーバプログラム
112…ドキュメントデータ参照手段
113…ドキュメントデータ更新手段
114…データ更新トリガ送信手段
115…検索エージェントプログラム
116…データ更新トリガ受信手段
117…ドキュメント索引生成手段
118…ドキュメント索引転送手段
120…ドキュメントデータベース
130…CPU
140…ネットワークインタフェース
150…更新バッファ
151～153…更新レコード
200…WWWサーバ
210…プログラムメモリ
211…WWWサーバプログラム
212…HTMLデータ参照手段
215…検索エージェントプログラム
216…データ更新検知手段
217…ドキュメント索引生成手段
218…ドキュメント索引転送手段
220…HTMLデータファイル群
230…CPU
240…ネットワークインタフェース
250…バッファ群
252…ファイル一覧バッファ
254…前回ファイル一覧バッファ
256…更新バッファ
300…検索サーバ
310…プログラムメモリ
311…検索サーバプログラム
312…ドキュメント索引検索手段
313…ドキュメント索引更新手段
320…ドキュメント索引情報
330…CPU
340…ネットワークインタフェース
400…クライアント
410…プログラムメモリ
411…クライアントプログラム
430…CPU
440…ネットワークインタフェース
500, 520, 540, 560, 580…通信回線

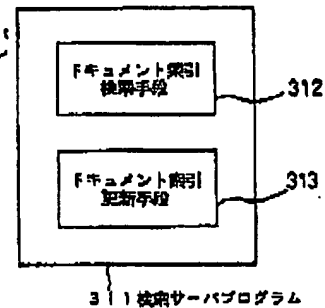
502...ドキュメント索引更新要求

504~508...転送レコード

【図1】

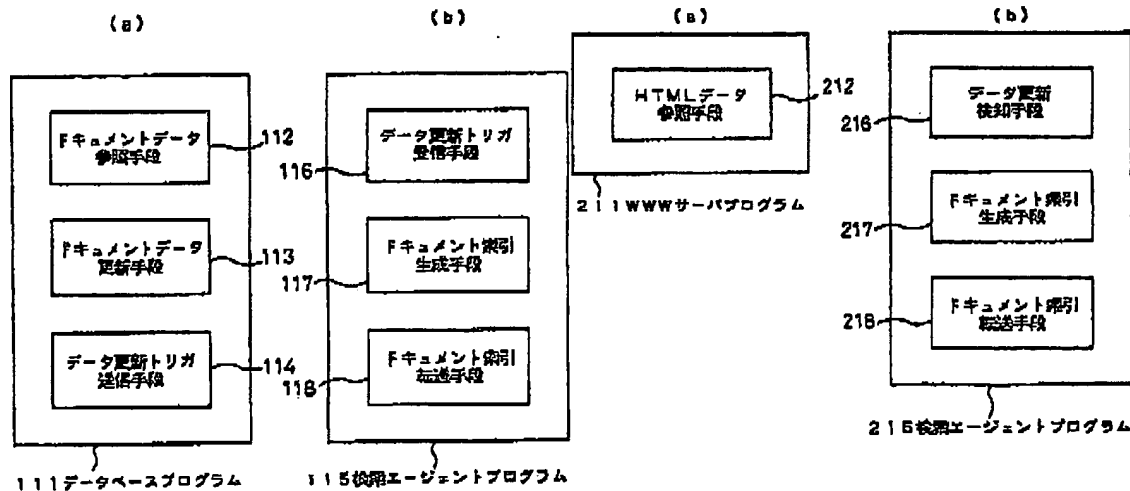


【図5】



【図3】

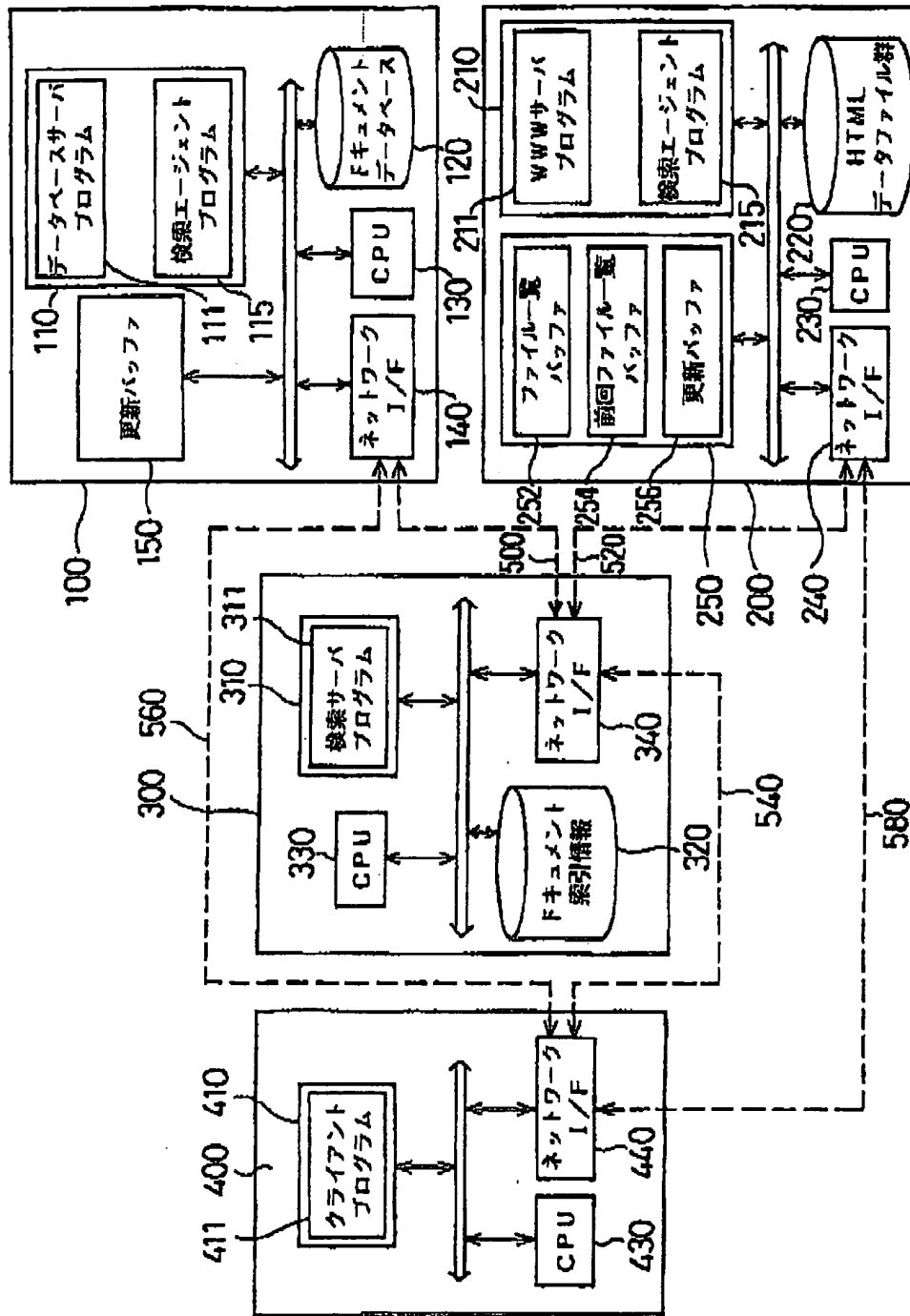
【図4】



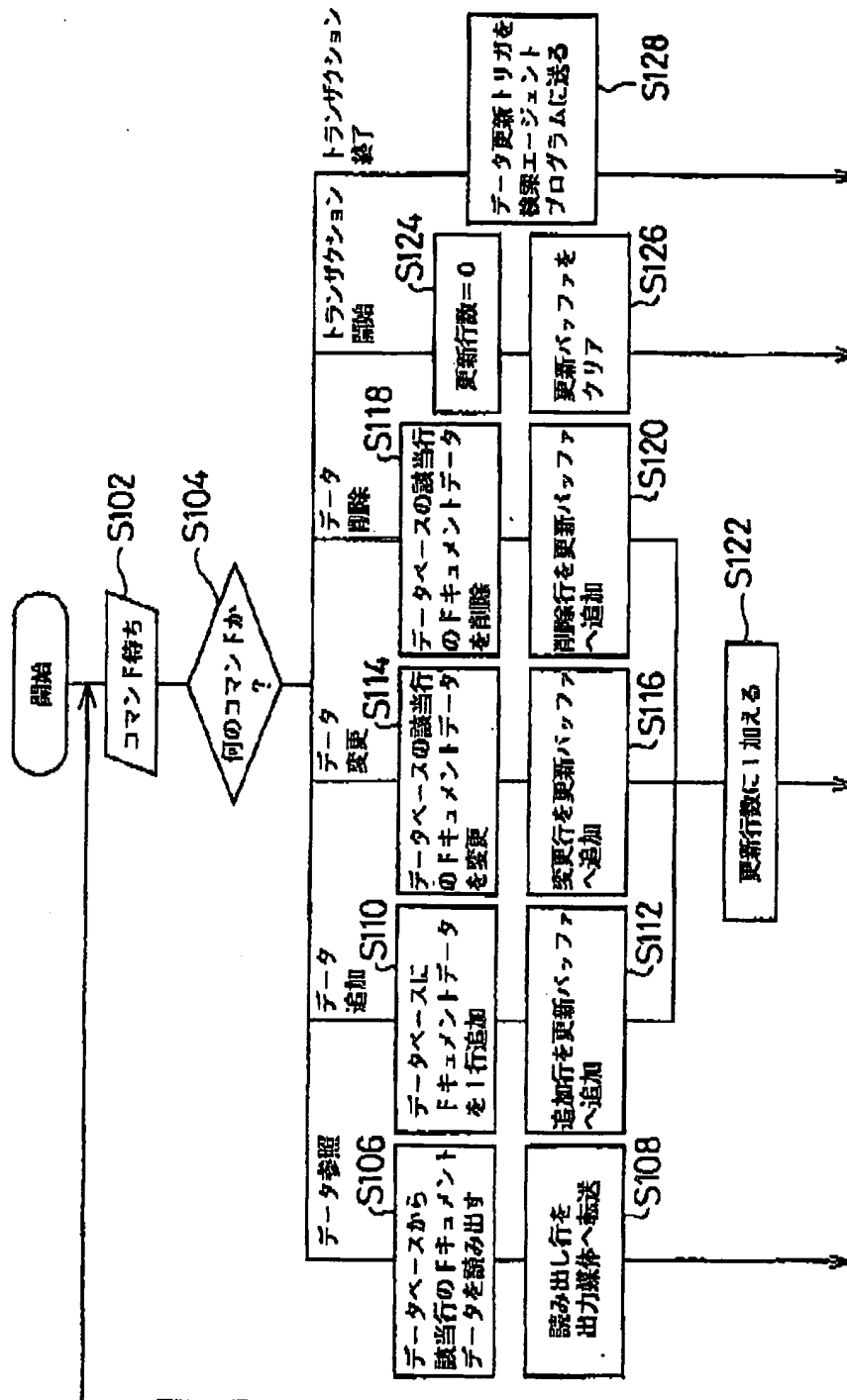
【図11】

| | | |
|--------------|-----------|-----|
| ドキュメント索引更新要求 | | 502 |
| 追加 | ドキュメント索引1 | 504 |
| 変更 | ドキュメント索引2 | 506 |
| 削除 | ドキュメント索引3 | 508 |
| . | . | . |
| . | . | . |
| . | . | . |
| . | . | . |

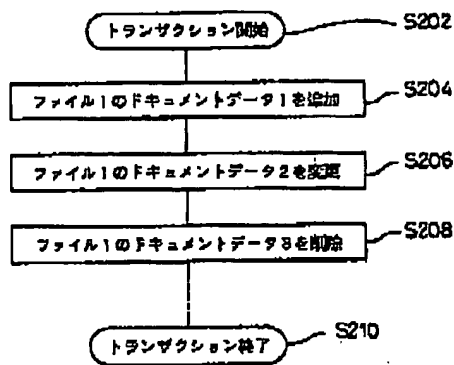
【図2】



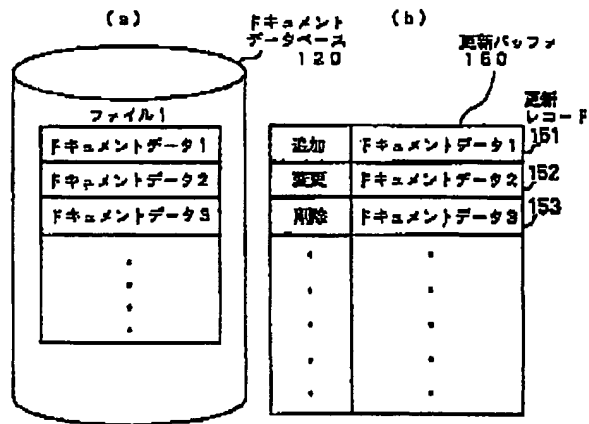
【図6】



【図7】



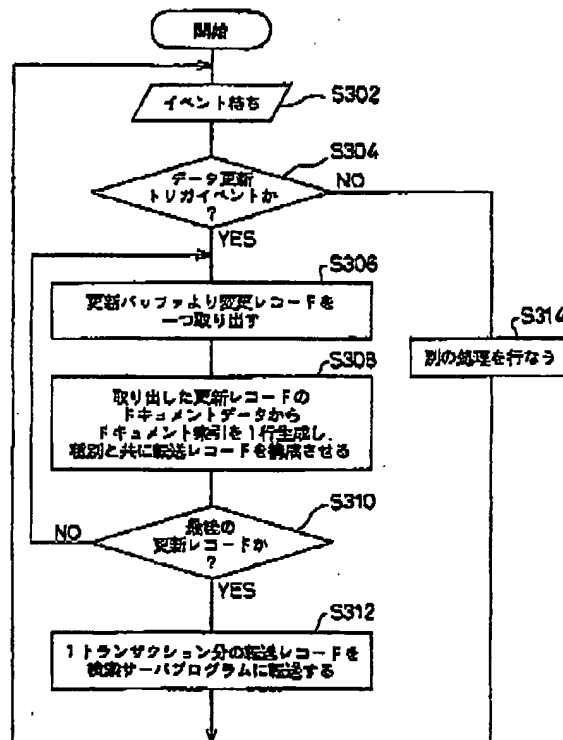
【図8】



【図9】

| |
|--------|
| Fコメント名 |
| 作者名 |
| 作成部門 |
| 作成年月日 |
| キーワード |
| 備考 |
| 本文 |

【図10】

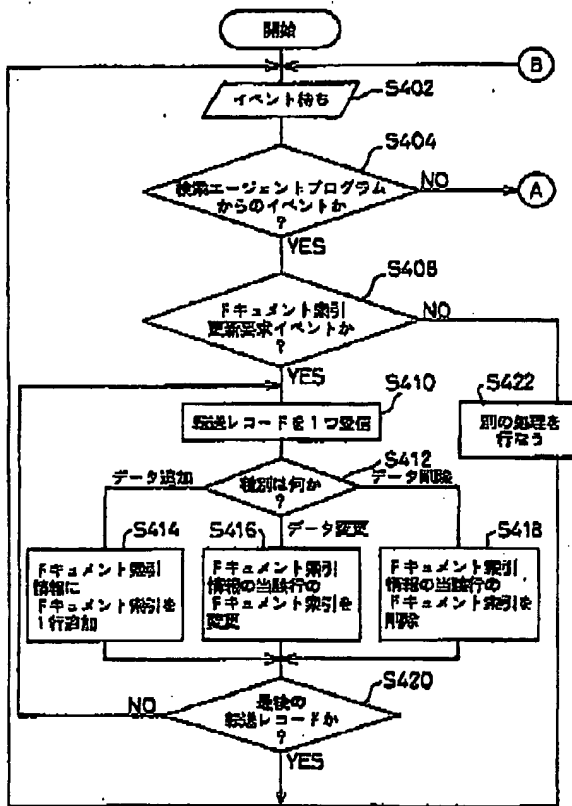


【図12】

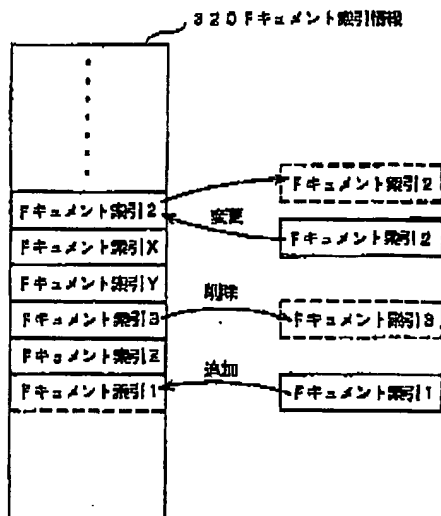
| |
|---------------|
| Fコメント名 |
| 作者名 |
| 作成部門 |
| 作成年月日 |
| ⋮ |
| キーワード |
| Fコメント種別 |
| Fコメントデータの格納場所 |

連結情報

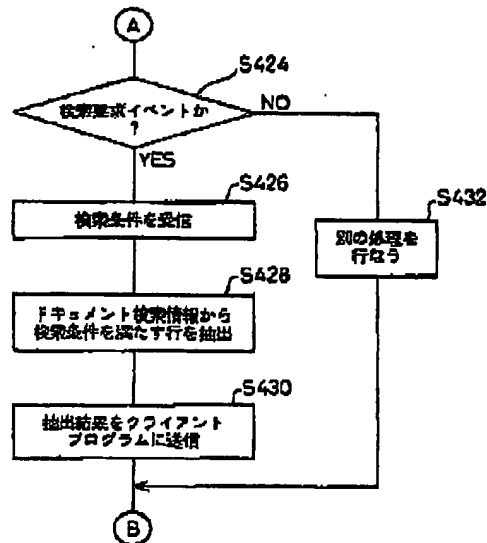
【図13】



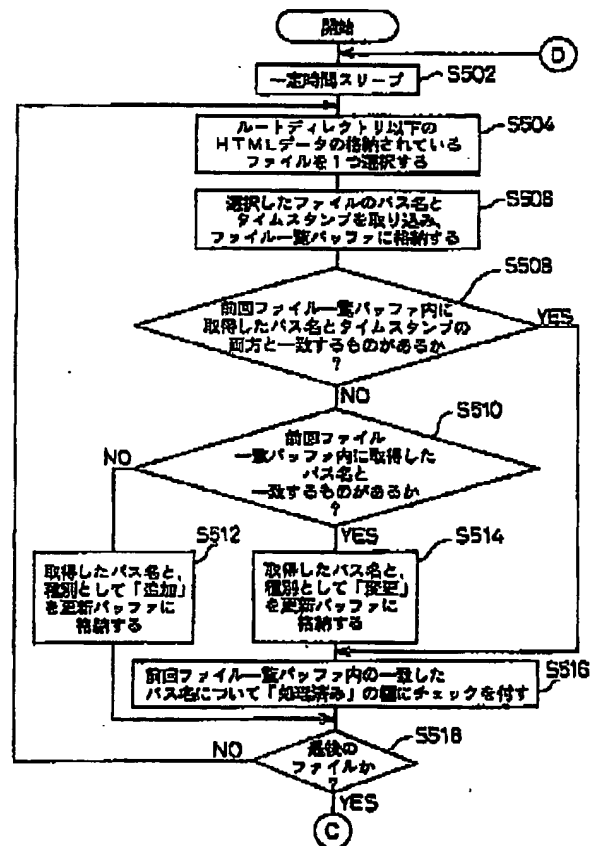
【図15】



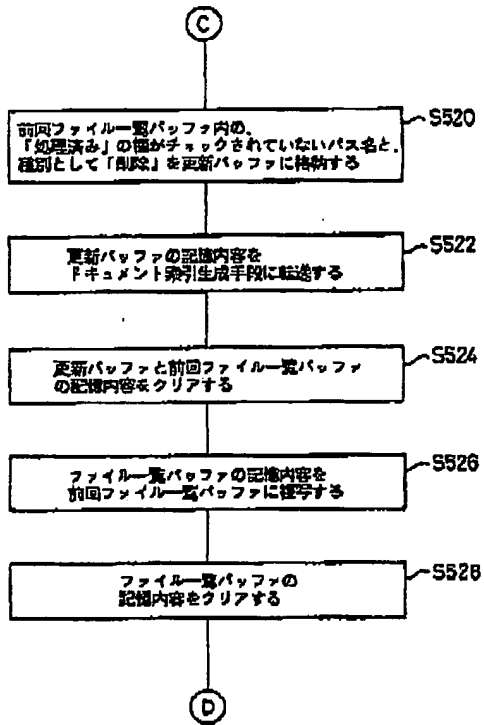
【図14】



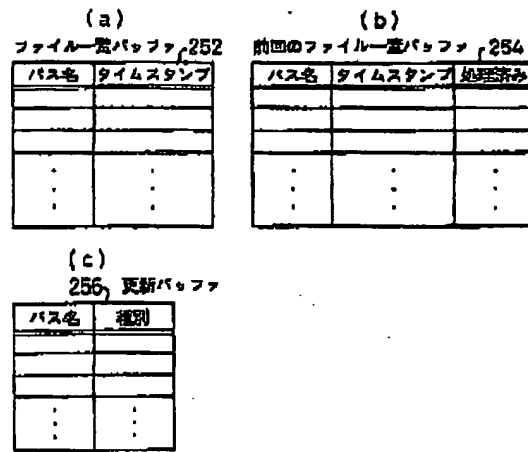
【図16】



【図 17】



【図 18】



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.